Separating between Unobserved Consumer Types: Evidence from Airlines*

Diego Escobari† Manuel A. Hernandez‡

September 25, 2018

Abstract

We propose an alternative approach to identify unobserved consumer types and assess whether firms price discriminate. Unlike other screening schemes that rely on quantity discounts or product differentiation, in our finite mixture structure individuals have unit demands and the product is homogeneous. We implement the model using an original U.S. airlines data set. The results support the existence of two demand types. The high type "business" traveler is less price sensitive, has a higher valuation and pays a higher price than the low type "tourist". The proportion of high types also increases as the departure date nears.

Keywords: Unobserved types, Price discrimination, Consumers' valuations, Airlines

JEL Classifications: C23, L93, R41

*We thank seminar participants at the International Industrial Organization Conference, Southern Economics Association, Latin American and Caribbean Economic Association and INESAD for helpful comments. We would also like to thank Wesley Wilson and two anonymous referees for their many useful comments. Stephanie C. Reynolds provided outstanding assistance with the data.

†Department of Economics & Finance, The University of Texas Rio Grande Valley, Edinburg, TX 78539, Phone: (956) 665-3366, Fax: (956) 665-5020, Email: diego.escobari@utrgv.edu., URL: http://faculty.utrgv.edu/diego.escobari

‡Markets, Trade and Institutions Division, IFPRI, Washington DC 20006, Phone: (202) 862-5645, Fax: (202) 467-4439, Email: m.a.hernandez@cgiar.org., URL: https://sites.google.com/site/manuelhernandez27

1 Introduction

A necessary condition for a firm to price discriminate requires being able to separate consumers into different groups based on their differentiated demands for a good or service. The firm can then charge different prices to different groups or consumer types. The challenge to the seller is that consumers hold private information —the different consumer types are generally *unknown* to the seller. Price discrimination schemes based on mechanism design theory assume that the seller knows the general distribution of tastes and demands for their product and can exploit these differences by offering a menu of prices such that consumers reveal their private information (own type).[1] In practice, however, distinguishing between different consumer types is more difficult. Consumers do not only privately know their own types, but also the information associated with their type such as price sensitivity and willingness to pay. This paper uses an original airline data set to propose a new empirical approach that separates between unobserved consumer types under incomplete information.[2]

Separating between consumer types is generally easier with quantity discounts or different product characteristics. The seller can use the discounts or product attributes to offer a menu of price and quantity (quality) combinations to screen consumers. For example, an airline carrier might offer first class and economy class tickets to separate consumers with different willingness to pay. Our approach is different because the identification of types occurs under incomplete information where there is not a clear screening mechanism for sellers to separate between consumers. There are no quantity discounts —travelers have unit demands— and the product is homogeneous for all types —an economy class ticket. Moreover, our empirical approach is flexible enough to accommodate incomplete information coming from various sources mimicking a multidimensional screening problem (see, e.g., Armstrong, 1996 and Rochet and Choné, 1998).

The airline markets are well suited for our screening and price discrimination model with unobserved types for several reasons. Airline tickets are not transferable so arbitrage

[1] Mussa and Rosen (1978) and Maskin and Riley (1984) are early contributions to this literature.

[2] In the airline industry, carriers can certainly keep track of travelers' habits and preferences through their frequent flyer programs and purchase history, but still this is only partial information about consumers and they cannot, for example, charge different prices to different individuals.

opportunities do not exist. Moreover, airlines have well documented price dispersion and consumer heterogeneity. We can focus on monopoly routes to control for the effect of competition and on the least expensive economy class tickets to control for various ticket characteristics that are correlated with additional sources of price dispersion and consumer heterogeneity. Finally, our data covers sales and prices across a large number of flights and markets that also permits us to correlate the modeled consumer types with different observable characteristics at the flight and route level.

The results show that when consumers can be grouped into different demand types, the model with two types provides the best fit to the data. Moreover, we find that the identified high-type "business" traveler has a less price-sensitive demand, a higher valuation and pays a higher price than the low-type "tourist". We interpret the price differences as price discrimination.[3] We also find that the proportion of high-type consumers is lower but dominates during the days closer to the departure date. The estimated probability of observing a demand regime associated with high types is larger when fares are above the flight average and when most travelers in the aircraft have already made a purchase. These findings are consistent with the label of high types as "business" travelers. When extending the analysis to non-monopoly routes, we are also more likely to observe high-type consumers in routes connecting high-income cities, hub airports and in more concentrated markets.

The advantage of using a mixture specification to account for unobserved types is that the inclusion of an incomplete set of variables to explain individual heterogeneity may be sufficient to produce consistent estimates of differentiated demands (Gan et al., 2015; Henry et al., 2014).[4] Previous studies that use mixture densities to identify unobserved regimes include Lee and Porter (1984) to study cartel stability, Keane and Wolpin (1997) to model endowment heterogeneity in the career decision, Knittel and Stango (2003) to study the

[3] Our distinction between "tourist" and "business" travelers is intuitive and helps in the exposition. However, there are alternative definitions for these two types of travelers, for example, Dana (1999b) allows them to differ only in their disutility of flying. Moreover, as explained in Dana (1998), price discrimination is difficult to define satisfactorily, especially when there are alternative legitimate definitions of costs. We discuss this in section 5.

[4] Ultimately, the types modeled may account for unobservable characteristics beyond price sensitivities and consumer valuations.

facilitative power of focal points in collusion agreements, and Gan and Hernandez (2013) to examine if agglomerated hotels have a higher probability of following collusive regimes. Unlike these studies, we further provide evidence on the robustness of the mixture model proposed.

Our study ties into the extensive literature on price discrimination. Shepard (1991) identifies price discrimination in gas stations where self-service and full-service serve as the screening mechanism. Deneckere and McAfee (1996) present a model where intentionally damaging manufactured goods to price discriminate may result in a Pareto improvement. In a theoretical model that has implications for airline pricing, Dana (1998) shows that price discrimination can exist in the form of advance purchases when firms have no market power. In an oligopolistic setting, Liu and Serfes (2004) find that when information quality increases the number of identifiable consumer segments also increases. Clerides (2002) and Cohen (2008) consider the role of quality in the U.S. market for books and the paper towel market, respectively. Leslie (2004) studies the effect of quality in Broadway theaters and, along with mail discount coupons, quantifies the welfare effects of price discrimination. Coupons as a screening device are also considered in Nevo and Wolfram (2002) for ready-to-eat cereal products. Busse and Rysman (2005) focus on price discrimination in Yellow Pages advertising where sellers vary the size of the advertisement offered, while McManus (2007) uses quantity discounts to study product design efficiency in an oligopoly. More recently, Tappata and Cossa (2014) show that opaque bookings help hotels to segment their demand.

The study also helps to explain the widely documented price dispersion in airlines. While most of the previous work uses the Airline Origin and Destination Survey transaction data (DB1B) from the Bureau of Transportation Statistics, these data is too aggregate to analyze price discrimination. Using these data, e.g., Gerardi and Shapiro (2009), Dai et al. (2014), Dana and Orlov (2014), help explain price dispersion, while Sweeting et al. (2016) develop a dynamic limit pricing model to explain why incumbent airlines cut prices on routes threatened with entry by Southwest Airlines to deter entry. Posted prices from online travel agencies have been helpful to study pricing and consumer decisions in greater detail. McAfee and te Velde (2007), for example, study dynamic pricing, while Bilotkach and Rupp (2011) analyze price-offer curves. Moreover, Escobari (2012) shows that airlines dynamically adjust prices in response to aggregate demand learning, Alderighi et al. (2015) analyze the roles of days to departure and inventories on prices, and Bilotkach et al. (2015) study yield management intervention.

Specific empirical work on price discrimination in airlines includes Stavins (2001) who approximates price discrimination with marginal implicit prices of ticket restrictions and, more recently, Hernandez and Wiggins (2014) who analyze nonlinear pricing strategies based on ticket menus and Escobari and Jindapon (2014) who look at the refundability of tickets as a screening device to price discriminate. Moreover, Lazarev (2016) studies the welfare effects of intertemporal price discrimination, while Williams (2017) estimates a model of dynamic airline pricing that accounts for intertemporal price discrimination and dynamic adjustment to stochastic demand. We focus on a homogenous-product setting. In this sense, our work is also related to the theoretical model on price dispersion with homogeneous goods and perfect competition of Prescott (1975) and more formally developed in Eden (1990). Dana (1999a) extends Prescott's model to imperfect competition and monopoly, while Deneckere and Peck (2012) generalizes it to a dynamic multi-period setting.

Our proposed model is also related to Berry et al. (2006) and Berry and Jia (2010). Berry et al. (2006) present a model to separate the effects of airline hubs on costs and markups, and Berry and Jia (2010) estimate the impact of airline demand and supply shocks on profitability. While their main research questions are different, we all use a mixture structure to model passenger heterogeneity. Berry et al. (2006) and Berry and Jia (2010) estimate "discrete-type" versions of Berry, Levinsohn, and Pakes' (1995) random coefficients model, in which the number of consumer types is finite. Discretizing the types helps them to reduce the computational burden of the estimation.[5] There are four main differences in our approach. First, our identification of different consumer types comes from variation in sales and prices at different points prior to departure (information unavailable in their data), while their type-specific parameters are identified from the substitution patterns among similar products when the mix of products varies across markets. Second,

[5] Escobari (2017) estimates a continuous version of Berry, Levinsohn, and Pakes' (1995) random coefficient model for airlines using a similar data set as ours and reduces the computational burden by aggregating across products.

unlike their work, we begin with the selection of the number of types. Third, we capture consumers' heterogeneity through the types and correlate the probability of observing a particular type with a set of observable characteristics. Fourth, our approach allows us to make inferences on whether firms price discriminate.

The remainder of the paper is structured as follows. Section 2 summarizes the data while Section 3 presents the empirical approach used to account for different consumer types. The estimation results are presented and discussed in Section 4, first focusing on monopoly markets and then extending the sample. Section 5 concludes.

2 Data

We have an original panel data set that contains posted prices and inventories gathered from the online travel agency Expedia.com. All flights departed on a single day, Thursday, June 22, 2006 to control for likely differences in prices and consumer heterogeneity across different departure dates. Stavins (2001) also collected posted prices on a Thursday to avoid weekend travel. We record inventories (obtained from the seat availability maps) and prices every three days for 228 flights during 103 days prior to departure. Recording data every three days helps to have enough variation between observations in time, while we argue that 103 days prior departure should capture most of the selling season for domestic travelers.

The flights in the sample are from American, Alaska, Continental, Delta, United, and US Airways, in which the proportion of flights by carrier was selected to resemble their share in the U.S. market. The sample contains 81 monopolistic and non-monopolistic routes (airport pairs). Selecting monopolistic routes helps us to control for competition in our base model, while the inclusion of non-monopolistic routes with varying number of sellers helps us to assess the role of competition in the model extension. Tickets are one-way non-stop to account for potential price differences associated with more sophisticated itineraries and potential consumer differences (e.g., in a round trip-ticket the length-of-stay and a Saturday-night stay might affect the consumer demand). We also focus on non-refundable economy-class tickets to control for more expensive first-class tickets and refundable tickets. We assume those tickets are of a significantly different quality.

The seat availability maps from Expedia.com show the total number of seats as well as available seats in the aircraft. When a sale occurs the map changes to reflect one less available seat. Because these maps showed seats only as available or as occupied, we do not know if the airline might decide to block available seats and hide them as unavailable. Seats might be blocked when the seat is broken, due to crew rest, weight and balance, or the seller might reserve seats for handicapped passengers until the day of departure (Williams, 2017). Unless they are repaired during the selling season, broken seats are a time-invariant characteristic, and hence controlled for. Handicapped passengers are paying passengers so those are likely recorded as regular sales. However, we do not know the number of crew-member blocked seats that might show as occupied, and it is not immediately clear how these blocked seats would impact the results. We argue it is reasonable that changes in those blocked seats represent only a small fraction in our sample.

Figure 1 presents the average fare for each of the 96 flights in monopolistic routes along with its corresponding +/− 1.96 within-flight standard deviations. Flights are ordered based on the observed within-flight price dispersion, as measured by the standard deviation of prices. We focus on monopolies to isolate the potential effect of competition on prices. The figure shows important within-flight price dispersion even across tickets for the same product. Given that the product is homogeneous, our study of how airlines segment consumers and price discriminate is not via self-selection schemes as in Escobari and Jindapon (2014).[6]

[Figure 1, here.]

Table 1 reports the summary statistics of the variables used in the analysis. We distinguish between monopolistic routes (48 airport pairs), which constitute the basis of our study, and the full routes sample. Fare is the posted price and expected, the average fare in monopoly routes is higher than in the full sample (323 versus 292 dollars or 41 versus 35 cents per mile when accounting for distance). Days is the number of days in advance fares and inventories are recorded and Load, our measure of inventories, is calculated as the ratio of occupied seats to total number of seats in the aircraft, i.e., it ranges from zero if the plane is empty to one if the plane is full. Sales (Q) are calculated as the difference

[6] Escobari and Jindapon (2014) show that airlines use the refundability of the ticket as a screening device.

between end and beginning-of-period inventories, $\text{SALES}_t = \text{LOAD}_{t+1} - \text{LOAD}_t$.[7] Once we record a sale, we assume that it occurred at the beginning-of-period one-way posted price.

[Table 1, here.]

Note that some passengers might be purchasing the observed segment paying a refundable fare or as part of a one-stop, multi-stop or round-trip itinerary. We argue that the observed one-way fares are relevant to our analysis as they are the base for the prices of other tickets that offer the same available seat. The simplest example to illustrate this point follows from Borenstein and Rose (1994) and Gerardi and Shapiro (2009) who assume that round-trip fares are just one-way fares multiplied by two. In this case, one-way fares perfectly correlate with round-trip fares. Likewise, Bachis and Piga (2011) explain that European Low Cost carriers price each segment independently such that the final price of the ticket is just the summation of the prices of each of the segments.[8] These two examples show that our observed one-way fares serve as basis for other more sophisticated itineraries that sell the same seat. Escobari (2012, 2017) and Williams (2017) follow a similar argument when recording fares and sales.

We also include two indicator variables that change by flight and over time to help us model the probability of observing a demand regime associated with a particular consumer type. These dummy variables indicate if FARE at a particular point in time is above the flight average ($\text{I}_{\text{FARE}>\overline{\text{FARE}}}$) and if LOAD is above the flight average ($\text{I}_{\text{LOAD}>\overline{\text{LOAD}}}$). When considering the full sample of routes, we can further associate the probability of observing a particular demand type with different route characteristics. INCOME is the average median household income in U.S. dollars between the departing and arrival cities. LEISURE is a dummy variable equal to one if the departing or arrival airport is located in Las Vegas or Orlando, zero otherwise. SLOT and HUB are also dummy variables. The first equals one if the number of landings and takeoffs in either airport are regulated, zero otherwise, while the latter is equal to one if the carrier has a hub in the origin or destination airport,

[7] For example, $Q = 0.02$ is equivalent to two seats sold in a 100-seat aircraft. The data contained a few observations in which the available seats increased. This might occur due to changed or canceled tickets. The results are not affected by this small fraction of negative sales.

[8] For the relationship between refundable and non-refunable fares, see Escobari and Jindapon (2014).

zero otherwise.[9] DISTANCE is the distance in miles between the two endpoint airports in a route while HHI is the Herfindahl-Hirshman Index that measures the concentration on the route. The index is constructed based on the number of seats in direct flights offered by each carrier on the airport pair.

3 Empirical Model

In this section we develop an empirical model to address the potential omitted-variable problem in the estimation of differentiated airline demands with unobserved consumer types. The key source of asymmetric information are the consumers' valuation for a good or service. In the case of perfect information sellers can first-degree or perfectly price discriminate. In a limited information setting, any information that correlates or permits to uncover consumers' valuations, at least partially, is potentially valuable. The result is a third-degree discrimination scenario.

The estimation uses a mixture model to explicitly account for the limited information on consumer types when estimating the demand for air travel. Consider a model in which the seller starts posting the price P, then based on this price consumers decide to buy the good or leave the market. This is consistent with Williams (2017), who shows that increasing prices over time provides little incentives for consumers to wait to purchase later. We model the existence of N different consumer types who have differentiated demands given by,

$$Q_{ijt} = \begin{cases} \alpha_1 + \beta_1 P_{ijt} + X\delta_1 + \kappa_i + \varepsilon_{ijt,1} & \text{if } \theta = 1, \\ \alpha_2 + \beta_2 P_{ijt} + X\delta_2 + \kappa_i + \varepsilon_{ijt,2} & \text{if } \theta = 2, \\ \vdots & \vdots \\ \alpha_N + \beta_N P_{ijt} + X\delta_N + \kappa_i + \varepsilon_{ijt,N} & \text{if } \theta = N, \end{cases} \tag{1}$$

where Q_{ijt} is sales for flight i on route j during period t. X is the matrix of observable explanatory variables that serve as controls, κ_i are flight fixed effects that account for time-invariant observed and unobserved differences across flights, and ε_{ijt} is an error term.

In this model, all factors associated with the unobserved demand types are absorbed by the constant terms α_θ for $\theta = 1, ..., N$. The different coefficients across types further

[9] The slot-controlled airports include Washington-National (DCA), New York-Kennedy (JFK), New York-La Guardia (LGA), and Chicago-O'Hare (ORD).

permit to capture differentiated effects of prices (P) and other factors (X) on sales by demand type.[10] For example, the degree of price sensitivity in the demand equation may vary across types.

If we assume that the error terms in Equations (1) are normally distributed such that $\varepsilon_{ijt,\theta} \sim N(0, \sigma^2_{\varepsilon,\theta})$, $\theta = 1, ..., N$, the log-likelihood for the kth flight-time period is given by,

$$\ln l_k = \ln \left[\sum_{\theta=1}^{N} \frac{r_\theta}{\sigma_{\varepsilon,\theta}\sqrt{2\pi}} \exp\left(\frac{-\varepsilon^2_{k,\theta}}{2\sigma^2_{\varepsilon,\theta}} \right) \right] \tag{2}$$

where r_θ is the mixing parameter defined as the probability of being in a regime of type θ consumers and $\sum_{\theta=1}^{N} r_\theta = 1$. Hence, each kth observation can be associated to a particular demand regime θ, $\theta = 1, ..., N$, with probability r_θ. Alternatively, we can think of multiple types coexisting in each observed period where the type-mixture can change over time, reflected in the estimated probabilities r_θ.

As we allow for varying coefficients across types, it is also of interest to parameterize the probability of observing a particular demand type. More specifically, we can model the probability of observing type-θ demand as,

$$r_\theta = \frac{\exp\left(G\delta_\theta\right)}{1 + \sum\limits_{s=1}^{N-1} \exp\left(G\delta_s\right)} \tag{3}$$

where G is a set of observable characteristics. We can accordingly associate the identified consumer types to specific characteristics. For example, G can contain the number of days prior to departure, indicator variables for periods where the fare is above the flight's average fare or if the load factor is above the flight's average, as well as different route controls (e.g., market concentration or other market characteristics). Note that in this setup the variables included in G do not determine or induce a ticket purchase, i.e., are not directly associated with the flight sales Q modeled in Equations (1). By conditioning on the demand type, any relationship between G and Q is solely driven by the relationship between G and the probability of observing a specific demand type r_θ. Refer to the model identification section in the Appendix for further discussion.

[10] This flexibility in the mixture structure is similar to Gan and Hernandez (2013). It also resembles allowing for different parameter sets by type in a random coefficient demand framework (where consumer heterogeneity is accommodated in this case by allowing the taste parameters to vary with individual characteristics).

One concern in the estimation of the model above is the potential endogeneity of prices (P) in the sale equations in (1). Fares may be endogenous both because of the presence of unobservable factors affecting both prices and sales and because P and Q may be jointly determined. However, as discussed earlier, the type component is absorbed by the constant terms in Equations (1), which permits to assume away any correlation between the unobserved types and prices, thereby accounting for this potential source of endogeneity. Similarly, we estimate the model in mean deviations to control for unobserved flight-specific characteristics that are time-invariant, such that the κ_i term is dropped out from (1).[11] Lastly, the use of posted prices (as opposed to aggregate transaction data) reduces the possibility that P and Q are jointly determined. In particular, consistent with Deneckere and Peck (2012) where firms start posting prices and then consumers arrive to observe posted prices and decide whether to purchase, we have posted prices that are set at the beginning of the period (at t). Then the modeled consumers' response to these posted prices (sales) is captured as the difference in load factors at the end of period $(t + 1)$ versus the beginning of period (t); hence, price is less likely to be endogeneous than if we were working with transaction data in which aggregation into periods results in a joint determination of demand and prices. Certainly, prices may still be endogenous if sellers are forward-looking and anticipate a demand shock before posting prices, reason why we empirically test for the endogeneity of P in the Appendix. The results summarized in Table A1 find no evidence supporting the endogeneity of this variable.[12]

4 Results

4.1 Pooled Demand

We now turn to the estimation results of the demand (sales) equations. To control for any effect of competition we focus on monopoly routes. We define a monopoly route as a

[11] We still include a constant term in the estimations in mean deviations, which could be viewed as a deterministic change or trend in sales.

[12] Note that capacity could also be viewed as endogenous as airlines can potentially change the aircraft size during the selling season in response to large demand shocks. During the data collection we did not observe though any aircraft size changes. Hence, capacity is time-invariant and controlled for as the model is estimated in mean deviations.

route in which there is only a single operating carrier. Note that this is a stricter definition than the one used in Borenstein and Rose (1994), who define a monopoly route when a single carrier operates more than 90% of the weekly direct flights. We first estimate a demand model in which we assume all consumers are of a single type. The first column of Table 2 (Model 1) shows the MLE results when pooling across all observations. For clarity of exposition, the dependent variable, sales (Q), is pre-multiplied by 100.

[Table 2, here.]

In Model 1 with no varying demand (consumer) types, we find that sales or changes in the load factor are negatively correlated with prices and with days prior to departure. The direction of these correlations are expected as we anticipate a downward-sloping demand for airline tickets as well as an increase in the number of purchases and consequent higher variation in the load factor as we approach the departure date. The coefficient of LnFare indicates that a 10% increase in prices is on average associated with a 0.237 decrease in seats in a 100-seat aircraft.[13] Regarding the effect of days prior to departure, the coefficient of Days shows that one day closer to departure is associated with an increase in sales of 0.033 seats for a 100-seat aircraft.

4.2 Differentiated Demands and Price Discrimination

The first step involves determining the number of types N. While the formulation of Equations (1) allow for a potentially large N, in practice we consider models between one and five types, i.e. $N = 1, ..., 5$, and select the number of types that best fit our data based on different selection criteria.[14] Table 3 reports the Schwarz Bayesian Information criterion (SBIC), the maximized value of the log likelihood function and the corresponding Likelihood Ratio (LR) tests comparing the fitness across models. We find that the resulting number of types is two ($N = 2$). The model with two types shows the lowest SBIC. Similarly, the

[13]The decrease in sales or change in the load factor ΔLoad is equal to $[2.366/(100 \times 100)] \times \%\Delta$Fare $= [2.366/(100 \times 100)] \times 10 = 0.002366$, which in a 100-seat aircraft is equal to 0.2366.

[14]Liu and Serfes (2004) show that the higher the quality of information on consumers' valuations available to a firm, the easier is for the firm to distinguish across different consumer types and price discriminate. While in the context of perfect information we can think of a continuum of types, under limited information it is more reasonable to consider a discrete (and limited) number of types.

LR test indicates that the two-type model provides a better fit than the one-type or pooled model, while the three-type model does not provide a better fit than the two-type model. Hereafter, we refer to the first consumer type as type H (high) and to the second consumer type as type L (low). In the model, each observed time period for a given flight can be linked to a consumer type, H or L, with a specific probability. Thus, within a flight the proportion of demand types can vary as we approach the departure date.

[Table 3, here.]

Several interesting results emerge when moving in Table 2 to Models 2 and 3 where we allow for differentiated demands by type (the difference between these two models is the variable included in the type equation discussed below). First, the estimation results clearly support the existence of heterogeneous consumers that can be separated into two types. Sales are more sensitive to prices in the demand associated with type-L individuals than in the demand associated with type-H individuals. In particular, $|\beta_L| > |\beta_H|$ suggesting that type-L individuals are more price-sensitive consumers, likely "tourists", while type-H individuals are less price-sensitive consumers, likely "business" travelers. The coefficient of LnFare is 1.8-2.4 times larger in the type-L demand compared to the type-H demand and the difference is statistically significant.[15] A 10% increase in prices decreases sales by 0.131-0.146 seats in a 100-seat aircraft when facing type-H travelers and by 0.264-0.316 seats when facing type-L travelers. The demand characterized by type-L travelers also exhibits a much higher dispersion in sales than the demand characterized by type-H travelers ($\sigma_{\varepsilon,L} > \sigma_{\varepsilon,H}$). The dispersion is roughly five times higher.

The estimations also permit to recover the likelihood of observing a demand with type-H travelers as denoted in Equation (3) for $N = 2$ and $\theta = H$. Model 2 specifies the mixing parameter r_H or probability of observing a demand with type-H consumers to be a function of days prior to departure ($\text{Days} - \overline{\text{Days}}$) while Model 3 includes an indicator variable for whether fares are above or below the flight's average fare ($\text{I}_{\text{Fare}>\overline{\text{Fare}}}$). Imposing a structure to the probability function allows us to associate the identified demand (consumer) types, i.e. type-H and type-L consumers, to specific observable characteristics. This also helps

[15] The corresponding Wald F statistic is 349.60 in Model 1 and 133.45 in Model 2, both with a p-value of zero. The difference in the price sensitivity between the two types holds across all the estimated models.

to reduce the possibility of alternative explanations for the results obtained and that the estimated model is not simply identifying two demand types by construction (or spuriously). We further discuss and formally test the identification of the proposed discrete mixture specification in the Appendix. The test results reported in Table A2 support the robustness of the estimated two-type model.

Model 2 reveals that we are more likely to observe type-H consumers as the departure date approaches. One day closer to departure increases the probability of observing type-H consumers by 0.4 percentage points. Model 3, in turn, shows that we are more likely to observe type-H consumers during periods when the fare is above the average fare in a given flight. In particular, when the fare is higher than the flight's average fare, the probability of observing a type-H demand is 18 percentage points higher than when the fare is equal to or below the flight's average fare (68% versus 50%). These patterns are in line with the notion that type-H consumers are more likely to be "business" travelers arriving closer to the departure date and buying at higher prices. While in Model 2 the type probability only varies over time, in Model 3 it varies over time and across flights. This is the reason why the latter specification is preferred to capture greater (or lower) relative presence of certain types of consumers as it can be associated with factors beyond the departure date.

From the estimation results we can also assess whether airlines price discriminate in the context of unobserved consumer types. In particular, three conditions must be met to conclude that price discrimination exists: (1) there is no arbitrage, (2) there are different consumer types with different demand preferences, and (3) different types should be paying different prices. The no-arbitrage condition is easily met as there is no ticket reselling among buyers. Second, the estimation results support the existence of two consumer types with different demands: a type-H demand with a lower price-elasticity and a type-L demand with a higher price elasticity. Third, we calculate that the average price paid by type-H consumers is higher than for type-L consumers. More specifically, from the estimated probabilities of observing a type-H demand in Model 3, we can separate the sample between observations that are more likely associated with type-H and type-L consumers.[16] We

[16]The estimated probabilities for each flight-time period observation are converted to a binary demand prediction (i.e. type-H or type-L) assuming that all observations with an estimated probability higher than the sample average are type-H demand periods and the remaining are type-L demand periods.

obtain that 1,133 observations belong to type-H demand periods and 2,110 observations to type-L demand periods and the corresponding sample-average prices are 381.6 dollars and 290.9 dollars. A larger number of type-L observations is further consistent with having a larger share of tourists.

From Model 3, we can additionally identify the proportion of each type at every point prior to departure, as Figure 2 illustrates. Consistent with type-H being "business" travelers and with the findings in Model 2, we observe that as the flight date nears the number of "business" travelers generally increases while the number of "tourists" decreases.

[Figure 2, here.]

4.3 Recovering Valuations

We can introduce a simple demand structure that allows us to recover the reservation values of each type. For notational convenience we drop the subscripts ijt. Let ω_θ denote the number of type-θ consumers, $\theta = L, H$. Reservation values for homogeneous airline seats are uniformly distributed $[0, \bar{v}_\theta]$, where $\bar{v}_\theta$ is the highest valuation for a seat. Hence the demand can be written as $Q = \omega_\theta - \omega_\theta/\bar{v}_\theta P$ and the number of consumers of each type is $\omega_\theta = \alpha_\theta + X\delta_\theta$. Then, the reservation values can be calculated as,

$$\bar{v}_\theta = -\frac{\alpha_\theta + X\delta_\theta}{\beta_\theta}. \tag{4}$$

Equation (4) shows that with estimates of $(\alpha_\theta, \beta_\theta, \delta_\theta)$ we can obtain $\bar{v}_H$ and $\bar{v}_L$ for particular values of X (days prior to departure).

To implement Equation (4) we need to estimate Equations (1) with fares in levels. The results of this alternative model, reported in Table 4, also support the existence of two consumer types. The estimates are very similar to those in Table 2. We observe a group of less price sensitive type-H individuals that can be labeled as "business" travelers and a group of more price sensitive type-L individuals that can be labeled as "tourists". Models 2 and 3 show that a 30-dollar increase in prices (roughly equivalent to a 10% increase in prices relative to the sample average) decreases type-H purchases by 0.129-0.138 seats in a 100-seat aircraft and decreases type-L purchases by 0.228-0.276.[17] These point estimates

[17]The decrease in the load factor ΔLOAD in Model 2 is equal to $0.0043 \times 30 = 0.00129$ among type-H

indicate that type-L consumers are roughly 65-114% more price sensitive than type-H consumers. We are also more likely to observe type-H travelers when the departure date is closer and when the fare is higher than the flight's average. Overall, these results show that our findings are not sensitive to the functional form of prices in the estimated demand model.

[Table 3, here.]

Combining Equation (4) and the estimates in Table 4 we derive the highest reservation values for type-H and type-L consumers. Because the model is estimated in mean deviations, the constant term used in Equation (4) is recovered as $\tilde{\alpha}_\theta = \bar{Q} - \hat{\beta}_\theta \bar{P} - \bar{X}\hat{\delta}_\theta$ for $\theta = H, L$. The lower part of Table 4 shows that type-H travelers exhibit a higher $\bar{v}$ than type-L travelers. On average, type-H travelers have a higher reservation value of 692.8-722.5 dollars versus 509.1-546.4 dollars of type-L travelers and the observed average fares are 348.6-381.6 dollars and 290.9-297.0 dollars, respectively. These results are consistent with airlines price discriminating and charging higher prices to individuals with higher valuations.

[Table 4, here.]

4.4 Alternative Model Specifications

Table 5 presents additional estimation results using alternative specifications for the probability r_H in Equation (3) that permits to associate the presence of type-H travelers with other observable characteristics, particularly with an indicator variable for periods when the load factor is higher than the average flight load factor. Model 1 includes the load factor indicator together with days prior to departure, while Model 2 further adds the indicator variable for fares above the flight's average. We find that travelers who arrive relatively late (i.e., after the average passenger in the flight already arrived, $\mathrm{I}_{\mathrm{LOAD} > \overline{\mathrm{LOAD}}} = 1$) are more likely to be type-H, which is consistent with "business" travelers labeled as the high type. Similarly, higher fares are associated with a higher probability of observing type-H

consumers and to $0.0092 \times 30 = 0.00276$ among type-L consumers, which in a 100-seat aircraft are equal to 0.129 and 0.276 seats; in Model 3 the decrease is equal to 0.138 and 0.228 seats.

travelers. An apparent counterintuitive result arises, however, with the positive sign on $\text{Days} - \overline{\text{Days}}$, which indicates that there is a lower probability of being type-H closer to the departure date. This is largely explained by the fact that most of the potential association between consumer types and days in advance is already captured by the indicator variables for the load factor (and to a lower extent prices) above the flight averages, which are highly correlated with the days prior to departure.

[Table 5, here.]

Based on the estimates in Model 2 we also construct Figure 3 to provide a clearer picture about the likelihood of observing a type-H demand. We find that we are more likely to observe type-H travelers when both the fare is above the average and when most of the travelers have already made a purchase, while we are more likely to observe type-L travelers when fares are low and before the average traveler in the flight has arrived.

[Figure 3, here.]

The results obtained in Table 5 and Figure 3 imply that the differences in prices paid by type-L and type-H consumers are also consistent with alternative explanations of price dispersion. For example, all else equal, when there are fewer seats in the aircraft the opportunity cost of a seat is higher. If there is a higher probability of observing a type-H consumer when there are less available seats, it is likely that the higher prices paid by type-H consumers are due to capacity-based pricing.

4.5 Extended Sample with Competitive Routes

In Table 6 we expand the sample size from 3,243 to 7,705 observations to include flights in non-monopoly routes. The goal is not to model consumer behavior in the presence of multiple sellers, but to assess if the results from monopoly routes obtained earlier can be extended to competitive environments in which passengers can switch their purchases to competing sellers. Models 1-3 replicate the models in Table 2 while Model 4 includes in G all three variables included in the previous monopoly specification. The magnitude of the estimated coefficients in the demand equation are generally similar to our base estimates and all previous results hold. There are clearly two consumer types where "tourist" travelers (type-L) have more price sensitive demands and appear in larger proportions. All the signs of the variables in G are also the same as before.

[Table 6, here.]

Working with a larger sample helps to correlate route specific characteristics with the probability of observing a type-H demand. Table 7 reports the results of three additional specifications for the probability r_H of observing type-H travelers. We add standard route characteristics that include average household income at the endpoints, if route involves a leisure endpoint (i.e., if one of the endpoints is Orlando or Las Vegas), if route involves a slot-controlled airport (i.e., if the number of landings and takeoffs in either the origin or destination airport is regulated), if one of the endpoints is a hub for the carrier, distance and the level of route concentration using the Herfindahl-Hirschman index (HHI) based on carrier seat shares. These variables provide variations across markets (routes) but not within observations in a given flight. Consistent across all three models, the higher the income, the larger the probability of observing "business" travelers. Moreover, Hub and HHI have both a positive association with r_H. These positive correlations can be explained by higher fares charged at hub airports as well as in more concentrated routes.

[Table 7, here.]

In sum, the estimated coefficients in the sales equation of both Tables 6 and 7 are very similar to those reported in Tables 2 and 4 and provide additional robustness to our base results. The alternative specifications for the mixing parameter r_H further support the notion that the type-H demand is characterized by less price-sensitive "business" travelers while the type-L demand is characterized by more price-sensitive "tourists".

5 Conclusion

In a context of no arbitrage, unit demands and perfect information, a firm can set different prices that match the willingness to pay of each buyer and extract all the consumer surplus. However, this scenario of perfect price discrimination is not feasible because consumers hold information that is unknown to the seller. As a response, sellers can use a variety of screening mechanisms, e.g., quantity discounts and product characteristics, to induce buyers to reveal private information. In this paper we propose using a mixture specification to separate between unobserved consumer types. Our approach does not rely on particular product attributes as a screening device but on partial information related to the consumer types, which can originate from multiple sources and is expected to be correlated with information unobserved to the seller (e.g., valuations, price sensitivity).

We implement our model using an original U.S. airline data set with prices and sales across a wide set of flights and monopoly markets. The estimation results support the existence of two consumer types. Consistent with the conventional knowledge on consumer heterogeneity in airline markets, the low types most closely resemble "tourists" while the high types can be regarded as "business" travelers. We find that high types are less price sensitive, have higher valuations and pay higher prices. We are also more likely to observe high types when the departure date nears (*timing* of arrival is important) as well as when most travelers have already booked a ticket (*order* of arrival is important).

The estimated differentiated demands are robust to alternative specifications of the type-probability equation, which supports the model identification, as well as to the inclusion of non-monopoly routes. In particular, we are more likely to observe high types in routes connecting high-income cities, hub airports and in less competitive routes. Overall, our study contributes to the understanding of price dispersion in airlines in a setting with homogenous goods as opposed to previous related studies. The proposed method could be easily extended to other industries in which sellers screen consumers to charge differentiated prices (e.g., fashion apparel, hotels, cruises, car rentals and the cable market for a specific product offered).

Lastly, we want to emphasize that whether the observed price differences between types constitute price discrimination is largely semantic (Dana, 1998). This is because the standard definition of price discrimination is based on differences in price markups over costs. Then, if costs differ the observed price differences are not discriminatory. In our setting and in Dana (1998), it is difficult to define the marginal costs. In particular, the shadow cost of a seat can vary over time in the period before the flight departs. Part of the reason fares are higher close to departure might be because flights fill-up and the shadow costs of capacity rises.

Appendix

Potential Endogeneity of Prices

One concern in the estimation of the sales equation in (1) is that prices (P) and the error term (ε) are potentially correlated. While the estimation accounts for the potential endogeneity that can arise from unobserved demand types and time-invariant flight-specific characteristics, another common source of the correlation between P and ε arises if the seller sets prices knowing more about the demand shock (ε) than the econometrician. Our data, however, is based on posted prices which means that sellers set prices first and consumers then respond to an already fixed posted price P.[18] Hence, P is less likely to be endogenous than if we were working with transaction data in which the aggregation into periods suggest that demand and price are jointly determined.

Still, the price might be endogenous if the seller is forward looking and able to anticipate a demand shock before posting its price. Hence, we perform a regression-based Hausman test to formally evaluate the potential endogeneity of fares in the estimated one-type sales equation. Due to the lack of an appropriate external instrument, we implement the test based on a model in first differences, which permits to rely on a sequential exogeneity assumption and use the lagged value of fares as an instrument (Anderson and Hsiao, 1981). In particular, $P_{ij,t-2}$ is a natural instrument for $\Delta P_{ij,t-1}$ since it is mathematically related to $\Delta P_{ij,t-1} = P_{ij,t-1} - P_{ij,t-2}$ but not to the error term $\Delta\epsilon_{ijt} = \varepsilon_{ijt} - \varepsilon_{ij,t-1}$. We naturally loose one observation per flight when following this approach.

[Table A1, here.]

As shown in the upper panel of Table A1, the first stage estimation results and corresponding under-identification and weak identification tests confirm the appropriateness of $P_{ij,t-2}$ as an instrument for $\Delta P_{ij,t-1}$. More important, the Hausman test reported in the bottom panel of the table does not reject the null hypothesis of exogeneity of fares in the estimated sales equation at conventional statistical levels. These results suggest that it is not necessary to rely on an instrumental variable approach to obtain unbiased estima-

[18] This is consistent with the theoretical model of posted prices in Deneckere and Peck (2012), where firms first post prices, then consumers arrive, observe the posted prices and decide whether to purchase.

tors. The results in Escobari (2012), who uses internal instruments validated via Sargan overidentifying restriction test, also support the exogeneity of fares.[19]

Model Identification

The estimation of a mixture model also requires evaluating the model identification. The identification of this class of models has been extensively studied in recent years (see, e.g., Mahajan, 2006, Gan et al., 2015, and Henry et al., 2014). On this matter, Henry et al. (2014) show that the model is fully identifiable if G is correlated with the unobserved regime type θ, i.e. that the probability of facing a type-H or type-L demand depends on G, and that G is conditionally independent of the error terms ε_H and ε_L in Equations (1).[20] Formally, the key identifying assumption in the proposed model requires $f\left(Q|\theta = H, P, X, G\right) = f\left(Q|\theta = H, P, X\right)$, where $f(\cdot)$ is the normal density function. Hence, conditional on facing a particular demand type, the factors that help to characterize θ are not related to the flight sales Q. That is, any association between G and flight sales is solely driven by the association between G and the probability of being in a certain demand type.

A direct implication of this key assumption is that we require some but not full information about the factors describing the unobserved demand types θ. When G includes more than one variable, Gan et al. (2015) and Henry et al. (2014) show that either using the full set of G or a subset of G will produce consistent estimates of the parameters in the sales equations in (1). We can implement a Hausman-type specification test to compare the estimated coefficients in Equations (1) using a full set of variables in G versus the estimates using a subset of variables in G.[21] Failing to reject the null hypothesis of no systematic differences between the estimated coefficients provides supporting evidence for the appropriateness of the model specification.

Table A2 reports the corresponding Hausman tests when using different subsets of variables in the demand-type probability equation. The benchmark model includes in the

[19]The estimates for fare in columns 2 (fare treated as exogenous) and 8 (fare treated as endogenous) in Table 4 are similar (see, Escobari, 2012).

[20]Intuitively, the model identification is similar to that underlying a two-stage least squares (2SLS) procedure.

[21]This test is similar to an overidentification test in an instrumental variables approach.

type equation indicator variables for periods where the fare and load factor are above the flight average ($\mathrm{I}_{\text{Fare}>\overline{\text{Fare}}}$ and $\mathrm{I}_{\text{Load}>\overline{\text{Load}}}$) as well as days in advance deviations from the flight average ($\text{Days} - \overline{\text{Days}}$). We compare this model versus the model that excludes $\mathrm{I}_{\text{Fare}>\overline{\text{Fare}}}$ and versus the model that excludes both $\mathrm{I}_{\text{Load}>\overline{\text{Load}}}$ and $\text{Days} - \overline{\text{Days}}$. We conclude that there are no systematic differences in the estimated coefficients of the sales equations across the models, which supports the robustness of the implemented mixture model.

[Table A2, here.]

References

Alderighi, M., Nicolini, M., and Piga, C. (2015). Combined effects of capacity and time on fares: Insights from the yield management of a low-cost airline. *Review of Economics and Statistics*, 97(4):900–915.

Anderson, T. and Hsiao, C. (1981). Estimation of dynamic models with error components. *Journal of the American Statistical Association*, 76(375):598–606.

Armstrong, M. (1996). Multiproduct nonlinear pricing. *Econometrica*, 64(1):51–75.

Bachis, E. and Piga, C. (2011). Low-cost airlines and online price dispersion. *International Journal of Industrial Organization*, 29(6):1655–667.

Berry, S., Carnall, M., and Spiller, P. (2006). *Airline hubs: Costs, markups and the implications of customer heterogeneity,* ed. Darin Lee, volume 1 of *Advances in Airline Economics: Competition Policy and Antitrust*, pages 183–214. Amsterdam: Elsevier.

Berry, S. and Jia, P. (2010). Tracing the woes: An empirical analysis of the airline industry. *American Economic Journal: Microeconomics*, 2(3):1–43.

Berry, S., Levinsohn, J., and Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica*, 63(4):841–890.

Bilotkach, V., Gaggero, A., and Piga, C. (2015). Airline pricing under different market conditions: Evidence from European low-cost carriers. *Tourism Management*, 47:152–163.

Bilotkach, V. and Rupp, N. (2011). A guide to booking airline tickets online. In Advances in Airline Economics: Pricing Behavior and Non-Price Characteristics in the Airline Industry, Vol. 3, 2012, ed. James Peoples.

Borenstein, S. and Rose, N. (1994). Competition and price dispersion in the U.S. airline industry. *Journal of Political Economy*, 102(4):653–683.

Busse, M. and Rysman, M. (2005). Competition and price discrimination in Yellow Pages advertisement. *Rand Journal of Economics*, 36(2):378–390.

Clerides, S. (2002). Book value: Intertemporal pricing and quality discrimination in the US market for books. *International Journal of Industrial Organization*, 20:1385–1408.

Cohen, A. (2008). Package size and price discrimination in the paper towel market. *International Journal of Industrial Organization*, 26:502–516.

Dai, M., Liu, Q., and Serfes, K. (2014). Is the effect of competition on price dispersion non-monotonic? Evidence from the U.S. airline industry. *Review of Economics and Statistics*, 96(2):161–170.

Dana, Jr., J. (1998). Advance-purchase discounts and price discrimination in competitive markets. *Journal of Political Economy*, 106(2):395–422.

Dana, Jr., J. (1999a). Equilibrium price dispersion under demand uncertainty: The roles of costly capacity and market structure. *Rand Journal of Economics*, 30(4):632–660.

Dana, Jr., J. (1999b). Using yield management to shift demand when the peak time is unknown. *Rand Journal of Economics*, 30(3):456–474.

Dana, Jr., J. and Orlov, E. (2014). Internet penetration and capacity utilization in the US airline industry. *American Economic Journal: Microeconomics*, 6(4):106–137.

Deneckere, R. and McAfee, R. (1996). Damaged goods. *Journal of Economics & Management Strategy*, 5(2):149–174.

Deneckere, R. and Peck, J. (2012). Dynamic competition with random demand and costless search: A theory of price posting. *Econometrica*, 80(3):1185–1247.

Eden, B. (1990). Marginal cost pricing when spot markets are complete. *Journal of Political Economy*, 98(6):1293–1306.

Escobari, D. (2012). Dynamic pricing, advance sales, and aggregate demand learning in airlines. *Journal of Industrial Economics*, 60(4):697–724.

Escobari, D. (2017). Airport, airline and departure time choice and substitution patterns: An empirical analysis. *Transportation Research Part A*, 103:198–210.

Escobari, D. and Jindapon, P. (2014). Price discrimination through refund contracts in airlines. *International Journal of Industrial Organization*, 34(3):1–8.

Gan, L. and Hernandez, M. (2013). Making friends with your neighbors? Agglomeration and tacit collusion in the lodging industry. *Review of Economics and Statistics*, 95(3):1002–1017.

Gan, L., Huang, F., and Mayer, A. (2015). A simple test for private information in insurance markets with heterogeneous insurance demand. *Economics Letters*, 136:197–200.

Gerardi, K. and Shapiro, A. (2009). Does competition reduce price dispersion? New evidence from the airline industry. *Journal of Political Economy*, 117(1):1–37.

Henry, M., Kitamura, Y., and Salanie, B. (2014). Partial identification of finite mixtures in econometric models. *Quantitative Economics*, 5(1):123–144.

Hernandez, M. and Wiggins, S. (2014). Nonlinear pricing strategies and competitive conditions in the airline industry. *Economic Inquiry*, 52(2):539–561.

Keane, M. and Wolpin, K. (1997). The career decisions of young men. *Journal of Political Economy*, 105(3):473–522.

Knittel, C. and Stango, V. (2003). Price ceilings as focal points for tacit collusion: Evidence from credit cards. *American Economic Review*, 93(5):1703–1729.

Lazarev, J. (2016). The welfare effects of intertemporal price discrimination: An empirical analysis of airline pricing in U.S. monopoly markets. Working Paper. New York University.

Lee, L.-F. and Porter, R. (1984). Switching regression models with imperfect sample separation information: Application on cartel stability. *Econometrica*, 52(2):391–418.

Leslie, P. (2004). Price discrimination in Broadway theater. *Rand Journal of Economics*, 35(3):520–541.

Liu, Q. and Serfes, K. (2004). Quality of information and oligopolistic price discrimination. *Journal of Economics & Management Strategy*, 13(4):617–702.

Mahajan, A. (2006). Identification and estimation of regression models with misclassification. *Econometrica*, 74(3):631–665.

Maskin, E. and Riley, J. (1984). Monopoly with incomplete information. *Rand Journal of Economics*, 15(2):171–196.

McAfee, R. and te Velde, V. (2007). *Dynamic pricing in the airline industry,* ed. Hendershott T.J., volume 1 of *Handbook of Economics and Information Systems.* Elsevier Science, New York, New York, U.S.A.

McManus, B. (2007). Nonlinear pricing in an oligopoly market: The case of specialty coffee. *Rand Journal of Economics*, 38(2):512–532.

Mussa, M. and Rosen, S. (1978). Monopoly and product quality. *Journal of Economic Theory*, 18:301–317.

Nevo, A. and Wolfram, C. (2002). Why do manufacturers issue coupons? An empirical analysis of breakfast cereals. *Rand Journal of Economics*, 33(2):319–339.

Prescott, E. (1975). Efficiency of the natural rate. *Journal of Political Economy*, 83(6):1229–1236.

Rochet, J.-C. and Choné, P. (1998). Ironing, sweeping, and multidimensional screening. *Econometrica*, 66(4):783–826.

Shepard, A. (1991). Price discrimination and retail configuration. *Journal of Political Economy*, 99(1):30–53.

Stavins, J. (2001). Price discrimination in the airline market: The effect of market concentration. *Review of Economics and Statistics*, 83(1):200–202.

Sweeting, A., Roberts, J., and Gedge, C. (2016). A model of dynamic limit pricing with an application to the airline industry. Working Paper. University of Maryland.

Tappata, M. and Cossa, R. (2014). Price discrimination 2.0: Opaque bookings in the hotel industry. Working Paper. University of British Columbia.

Williams, K. (2017). Dynamic airline pricing and seat availability. Cowles Foundation Discussion Paper No. 3003.

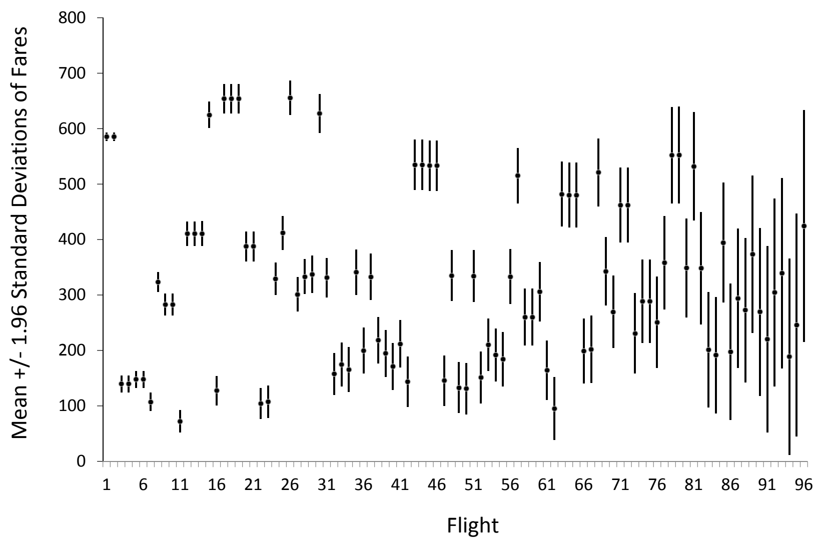

Figure 1: Mean +/- 1.96 Within-flight Standard Deviations of Fares (in Dollars)

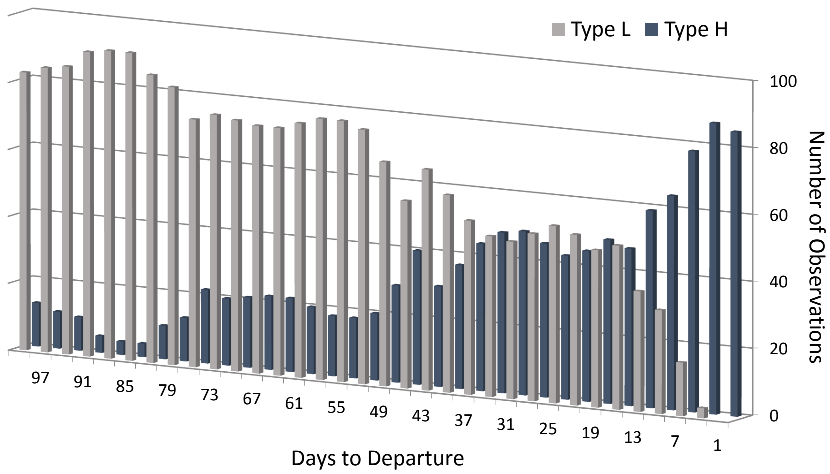

Figure 2: Number of observations by type and days in advance

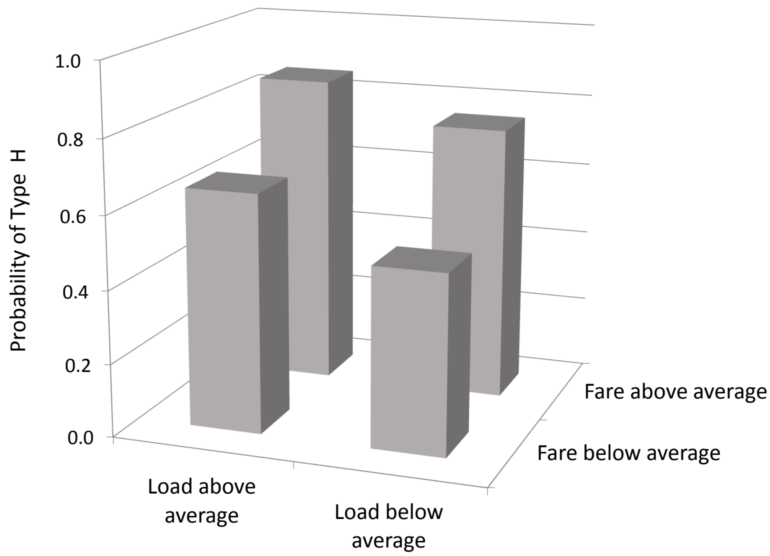

Figure 3: Probability of type-H demand

Table 1: Summary Statistics

VARIABLE	mean	sd	min	max
Monopolies:				
Fare (P)	322.6	171.7	64	914
Days	50.8	29.3	1	100
Load	0.544	0.245	0.013	1
Sales (Q)	0.017	0.038	-0.392	0.485
$I_{\text{Fare}>\overline{\text{Fare}}}$	0.349	0.477	0	1
$I_{\text{Load}>\overline{\text{Load}}}$	0.427	0.495	0	1
Full sample:				
Fare (P)	292.2	172.3	54	1,224
Days	50.8	29.3	1	100
Load	0.513	0.250	0.013	1
Sales (Q)	0.017	0.042	-0.408	0.485
Income	35,580.0	4619.4	25,198	53,430.0
Leisure	0.070	0.256	0	1
Slot	0.298	0.458	0	1
Hub	0.737	0.440	0	1
Distance	1104.4	620.7	91	2,604
HHI	0.679	0.289	0.253	1
$I_{\text{Fare}>\overline{\text{Fare}}}$	0.340	0.474	0	1
$I_{\text{Load}>\overline{\text{Load}}}$	0.416	0.493	0	1

Note: The number of observations is 3,243 for the monopolies and 7,705 for the full sample.

Table 2: Maximum Likelihood Estimates: Price Discrimination

Model:	(1)	(2)		(3)	
Type θ:	Pooled	H	L	H	L
Demand Equations:					
CONSTANT	1.6840*	-0.1484	2.9358*	0.0737	3.7413*
	(0.1496)	(0.0885)	(0.2821)	(0.0955)	(0.3788)
LNFARE	-2.3664*	-1.3063*	-3.1633*	-1.4578*	-2.6381*
	(0.3409)	(0.1448)	(0.6828)	(0.1581)	(0.9031)
DAYS	-0.0332*	-0.0153*	-0.0375*	-0.0175*	-0.0491*
	(0.0027)	(0.0013)	(0.0056)	(0.0014)	(0.0065)
σ_ε	3.7172*	0.9663*	5.2705*	1.1275*	5.5912*
	(0.0485)	(0.0267)	(0.1096)	(0.0380)	(0.1361)
Probability of Type H, $r_H = \text{Prob}(\theta = H)$:					
$\text{DAYS} - \overline{\text{DAYS}}$		-0.0145*			
		(0.0021)			
$\text{I}_{\text{FARE} > \overline{\text{FARE}}}$				0.7376*	
				(0.1138)	
Average Fare	322.6	348.6	297.0	381.6	290.9
Observations	3,243	3,243		3,243	
Log likelihood	6,075.1	7,103.3		7,106.9	
SBIC[a]	-3.737	-4.358		-4.360	

Note: The dependent variable is SALES $\times$ 100. Standard errors in parentheses. All regressions control for flight fixed effects. ‡ significant at 10%; † significant at 5%; * significant at 1%. [a] Schwarz Bayesian Information criterion.

Table 3: Determining the Number of Types N

Number of types N	SBIC[a]	Log likelihood	LR test[b]
1	-3.737	6075.057	
2	-4.360	7106.864	0.000
3	-4.351	7111.356	0.110
4	-4.340	7114.844	0.222
5	-4.331	7119.636	0.088

Note: [a] Schwarz Bayesian Information criterion. [b] p-value of likelihood ratio (LR) test reported.

Table 4: Maximum Likelihood Estimates: Reservation Values

Model:	(1)	(2)		(3)	
Type θ:	Pooled	H	L	H	L
Demand Equations:					
Constant	1.6529*	-0.1509	2.8429*	0.0787	3.6683*
	(0.1425)	(0.0968)	(0.2947)	(0.0917)	(0.3607)
Fare	-0.0074*	-0.0043*	-0.0092*	-0.0046*	-0.0076*
	(0.0010)	(0.0005)	(0.0021)	(0.0005)	(0.0025)
Days	-0.0326*	-0.0152*	-0.0357*	-0.0174*	-0.0475*
	(0.0025)	(0.0013)	(0.0057)	(0.0014)	(0.0063)
σ_ε	3.7190*	0.9663*	5.2775*	1.1356*	5.6126*
	(0.0475)	(0.0286)	(0.1066)	(0.0365)	(0.1326)
Probability of Type H, $r_H = \text{Prob}(\theta = H)$:					
$\text{Days} - \overline{\text{Days}}$		-0.0146*			
		(0.0020)			
$\text{I}_{\text{Fare} > \overline{\text{Fare}}}$				0.7602*	
				(0.1096)	
Reservation Values:	554.3	722.5	509.1	692.8	546.4
Average Fare	322.6	348.6	297.0	381.6	290.9
Observations	3,243	3,243		3,243	
Log likelihood	6,073.5	7,102.2		7,106.3	
SBIC[a]	-3.736	-4.358		-4.360	

Note: The dependent variable is Sales × 100. Standard errors in parentheses. All regressions control for flight fixed effects. ‡ significant at 10%; † significant at 5%; * significant at 1%. [a] Schwarz Bayesian Information criterion.

Table 5: Maximum Likelihood Estimates: Alternative Specifications of Type Equation

Model:	(1)		(2)	
Type θ:	H	L	H	L
Demand Equations:				
CONSTANT	0.0838	3.4119*	0.1491	3.5532*
	(0.1080)	(0.3692)	(0.1003)	(0.4205)
LNFARE	-1.4496*	-3.4422*	-1.4359*	-2.8008*
	(0.1972)	(0.9093)	(0.1878)	(1.0884)
DAYS	-0.0170*	-0.0431*	-0.0172*	-0.0436*
	(0.0016)	(0.0072)	(0.0015)	(0.0082)
σ_ε	1.1453*	5.7486*	1.2375*	6.0606*
	(0.0523)	(0.1687)	(0.0457)	(0.1901)
Probability of Type H, $r_H = \text{Prob}(\theta = H)$:				
$\text{I}_{\text{LOAD}>\overline{\text{LOAD}}}$	0.9974*		0.6630*	
	(0.1930)		(0.1687)	
$\text{DAYS} - \overline{\text{DAYS}}$	0.0261*		0.0303*	
	(0.0031)		(0.0032)	
$\text{I}_{\text{FARE}>\overline{\text{FARE}}}$			1.1533*	
			(0.1450)	
Observations	3,243		3,243	
Log likelihood	7,137.0		7,180.3	
SBIC[a]	-4.377		-4.401	

Note: The dependent variable is SALES $\times$ 100. Standard errors in parentheses. All regressions control for flight fixed effects. ‡ significant at 10%; † significant at 5%; * significant at 1%. [a] Schwarz Bayesian Information criterion.

Table 6: Maximum Likelihood Estimates: Extended Sample with Competitive Routes

Model:	(1)	(2)		(3)		(4)	
Type θ:	Pooled	H	L	H	L	H	L
Demand Equations:							
CONSTANT	1.8369*	-0.2858	3.0210*	0.0322	3.8670*	0.0632	3.5123*
	(0.1071)	(0.1614)	(0.2096)	(0.0664)	(0.2842)	(0.0733)	(0.2773)
LNFARE	-1.7219*	-0.8786*	-2.2825*	-1.0358*	-2.0037*	-0.9848*	-2.0272*
	(0.2048)	(0.0916)	(0.3890)	(0.1055)	(0.6350)	(0.1136)	(0.6096)
DAYS	-0.0362*	-0.0140*	-0.0392*	-0.0177*	-0.0522*	-0.0170*	-0.0443*
	(0.0018)	(0.0009)	(0.0042)	(0.0010)	(0.0050)	(0.0010)	(0.0057)
σ_ε	4.0316*	0.9371*	5.7375*	1.1038*	6.0516*	1.1661*	6.4194*
	(0.0343)	(0.0188)	(0.0755)	(0.0267)	(0.1003)	(0.0308)	(0.1381)
Probability of Type H, $r_H = \text{Prob}(\theta = H)$:							
$\text{DAYS} - \overline{\text{DAYS}}$		-0.0206*				0.0327*	
		(0.0013)				(0.0020)	
$\text{I}_{\text{FARE}>\overline{\text{FARE}}}$				0.5266*		0.9872*	
				(0.0750)		(0.0836)	
$\text{I}_{\text{LOAD}>\overline{\text{LOAD}}}$						0.5496*	
						(0.1112)	
Observations	7,705	7,705		7,705		7,705	
Log likelihood	13,807.7	16,396.5		16,471.5		16,713.5	
SBIC[a]	-3.579	-4.246		-4.265		-4.326	

Note: The dependent variable is SALES $\times$ 100. Standard errors in parentheses. All regressions control for flight fixed effects. ‡ significant at 10%; † significant at 5%; * significant at 1%. [a] Schwarz Bayesian Information criterion.

Table 7: Maximum Likelihood Estimates: Extended Sample with Competitive Routes and Alternative Specifications of Type Equation

Model:	(1)		(2)		(3)	
Type θ:	H	L	H	L	H	L
Demand Equations:						
CONSTANT	0.7170*	5.9771*	0.7208*	6.0161*	0.7209*	6.0107*
	(0.0670)	(0.5449)	(0.0712)	(0.6588)	(0.0680)	(0.6936)
LNFARE	-1.3195*	-2.8756*	-1.3209*	-2.8989*	-1.3235*	-2.8985†
	(0.1133)	(1.0855)	(0.1132)	(1.0175)	(0.1139)	(1.1793)
DAYS	-0.0239*	-0.0749*	-0.0239*	-0.0754*	-0.0239*	-0.0754*
	(0.0011)	(0.0103)	(0.0012)	(0.0109)	(0.0011)	(0.0124)
σ_ε	1.6183*	8.6465*	1.6228*	8.6719*	1.6220*	8.6680*
	(0.0343)	(0.2396)	(0.0325)	(0.2531)	(0.0319)	(0.2706)
Probability of Type H, $r_H = \text{Prob}(\theta = H)$:						
LNINCOME	0.1521*		0.1354*		0.1253†	
	(0.0062)		(0.0096)		(0.0589)	
LEISURE	-0.0695		-0.0004		-0.0855	
	(0.1742)		(0.1649)		(0.2126)	
SLOT			-0.0447		0.0569	
			(0.0928)		(0.1301)	
HUB			0.2679*		0.2663†	
			(0.0963)		(0.1095)	
LNDISTANCE					-0.0292	
					(0.0839)	
HHI					0.4188†	
					(0.1717)	
Observations	7,705		7,705		7,705	
Log likelihood	16,875.4		16,880.2		16,883.8	
SBIC[a]	-4.369		-4.368		-4.366	

Note: The dependent variable is SALES × 100. Standard errors in parentheses. All regressions control for flight fixed effects. ‡ significant at 10%; † significant at 5%; * significant at 1%. [a] Schwarz Bayesian Information criterion.

Table A1: Hausman Test for Potential Endogeneity of Fares

Dependent variable:	LnFare
First Stage Regressions:	
Lag LnFare	-0.0329*
	(0.0044)
Days	-0.0008*
	(0.0001)
Constant	0.2413*
	(0.0279)
Observations	3,145
Underidentification test:	
Kleibergen-Paap rk LM statistic	53.189
$\chi^2(1)$ P-val	0.000
Weak identification test:	
Kleibergen-Paap rk Wald F statistic	55.282
Hausman test. H_0: Fare is exogenous	
F(1,3141)	0.349
Prob > F(1,3141)	0.559

Note: Standard errors in parentheses. All regressions control for flight fixed effects. ‡ significant at 10%; † significant at 5%; * significant at 1%.

Table A2: Hausman Test to Evaluate the Model Identification

Excluded variables from type equation	H_0: Difference in coefficients of demand equation between benchmark model and alternative models not systematic
$\mathrm{I}_{\textsc{Fare}>\overline{\textsc{Fare}}}$	11.5929
	(0.1148)
$\mathrm{I}_{\textsc{Load}>\overline{\textsc{Load}}}$ & $\textsc{Days}-\overline{\textsc{Days}}$	9.4537
	(0.2217)

Note: Benchmark model includes $\mathrm{I}_{\textsc{Fare}>\overline{\textsc{Fare}}}$, $\mathrm{I}_{\textsc{Load}>\overline{\textsc{Load}}}$ and $\textsc{Days}-\overline{\textsc{Days}}$ in the type equation. Hausman Chi-squared statistic reported and p-value in parentheses.